IV. TECHNICAL NOTES

Cancer case definitions: A "cancer case" is defined as the primary cancer site, i.e., the site where the cancer started. Since an individual can have more than one primary cancer site, the number of incident cancer cases could be greater than the number of persons who are diagnosed with cancer. A metastasis is not a primary site.

Age-adjusted incidence rate: Age-adjusted incidence rates were calculated using the direct method, using 19 age groups-Census P25-1130 and standardized to the age distribution of the 2000 U.S. Standard Population (Appendix A). Age adjustment allows rates from one geographic area to be compared with rates from other geographic areas that may have differences in age distributions. Any observed differences in age-adjusted incidence rates between populations are not due to different age structures. Reports prior to 1999 used the 1970 U.S. Standard Population.

In confromity with the National Cancer Institute's (NCI) Surveillance, Epidemiology, and End Results (SEER) Program guidelines, the incidence rates for cancer sites exclude the following:

- ✓ *In situ* cases, except bladder;
- ✓ Basal and squamous cell skin cancers;
- ✓ Cases with unknown age; and
- ✓ Cases with unknown gender.

Age-specific incidence rates: Age specific rates are calculated by dividing the number of cases for a given age group by the total population of that age group and are expressed as an average annual rate per 100,000 persons by age group. Age specific rates exclude the same types of cases that are excluded from age-adjusted incidence rates. These rates, however, are crude rates, i.e. not age-adjusted.

Age-adjusted death rates: Death rates are calculated for total cases and separately for males and females using 19 age groups-Census P25-1130. The death rates are age-adjusted to the 2000 U.S. Standard Population (APPENDIX A). Rates are presented for 2003 and for the five -year period, 1999-2003.

Risks and associated risk factors: These were developed using the "American Cancer Society Textbook of Oncology, and the Harvard Cancer Center, *Causes of Human Cancer*.

Stage at time of diagnosis: Staging is the process of describing the extent or spread of disease from the origin, which is the primary site. Summary staging is the standard used for comparison nationally and is used throughout this report when references are made to stage at diagnosis. SEER Summary Stages 2000 are defined as follows:

In situ Malignant cells are within the cell group from which they arose, without penetration of the basement membrane of the tissue and no stromal invasion. *In situ* is "in place".

Localized The malignant cells are limited to the organ of origin and have spread no farther than the organ in which they started.

Regional The tumor is beyond the limits of the organ of origin by direct extension to adjacent areas with or without lymph node involvement.

Distant The primary tumor has broken away and has traveled, growing secondary tumors in other parts of the body. It has metastasized.

In situ and localized stages are the ***early stages*** of diagnosis. Regional and distant stages are ***late stage*** diagnoses. An **invasive cancer** refers to a cancer that has spread into surrounding tissues.

Years of life potential life lost (YPLL):

The years of potential life lost is a summary measure of premature mortality which provides an explicit way of weighting deaths occurring at younger ages, which are, a priori, preventable. The calculation of PYLL involves summing up deaths occurring at each age and multiplying this with the number of remaining years to live up to a selected age limit.

The **limit of 75 years** is currently used because the average life expectancy in the United States is over 75 years. $YPLL_{75}$ is calculated using the following eight age groups: under 1 year, 1-14 years, 15-24 years, 25-34 years, 35-44 years, 45-54 years, 55-64 years, 65-74 years. The number of deaths for each age group is multiplied by the years of life lost, calculated as the difference between age 75 years and the midpoint of the age group. For the eight age groups the midpoints are 0.5, 7.5, 19.5, 29.5, 39.5, 49.5, 59.5, and 69.5. For example, the death of a person 15-24 years of age counts as 55.5 years of life lost. Years of potential life lost is derived by summing years of life lost over all age groups.

Age-adjusted YPLL is calculated in order to assure trend comparison. The YPLL are age-adjusted and standardized, for each place i and each year t as follows:

$$PYLL_{it} = \sum_{a=0}^{l-1} (l-a)(d_{at} / p_{at})(P_a / P_n) * 100000$$

where *a* stands for age of the person a death , *l* is the upper age-limit, 75 years; d_{at} is the number of deaths at age *a*; p_{at} refers to the number of persons aged *a* in place *i* at time *t* ; P_a refers to the number of persons aged *a* in the reference population, and P_n refers to the total number of persons in the reference population.

2000 US Standard Population (YPLL age groups) Pa/Pn

00 years	0.01382
01-14 yrs	0.20087
15-24 years	0.13865
25-34 years	0.13557
35-44 years	0.16261
45-54 years	0.13483
55-64 years	0.08725
65-74 years	0.06604

SOURCE: Health, United States

Average years of life lost (AYLL): This is the extent to which life is cut short due to premature death. This is obtained by dividing the YPLL by the number of deaths. On average each person who dies from cancer loses 15 years of their life. The younger one dies, the more years are lost; therefore; death of children will have the highest AYLL numbers.

Confidence intervals (CI): A confidence interval tells how confident we are of the accuracy of the calculated rates. The SDCR uses a computed interval with a given probability of 95 percent, i.e., the true value of the calculated rate is contained within the interval. Thus, given a calculated rate of 191.4 and a confidence interval of 182.1 to 200.8, it is better to say that the true rate will fall between 182.1 and 200.8. The larger the sample size, the shorter the interval size, giving us more certainty that the rate is correct. When CI for percentages contains zero, the rate is considered to be stable. Above zero, it is the statistical significance is higher and below zero it is lower.

Mortality/incidence ratio (M/I): This ratio is calculated by dividing the number of deaths in a given year by the number of new cancers diagnosed in the same year. The death to case ratio provides a crude indication of the prognosis for patients. A ratio approaching 1.0, when the number of deaths equals the number of cases for a particular type of cancer, indicates a poor prognosis. A lower ratio indicates fewer deaths relative to the number of cases and suggests a better prognosis.

Statistical significance: This determines whether an event happens by chance alone. The null hypothesis states that in a given place and a period of time, all events occur randomly by chance. If not, then there is statistical significance. Confidence intervals are used to test statistical significance in this report. If the confidence intervals of two different rates intersect each other, then there is no statistical difference between the two rates. However, if the confidence intervals do not intersect one another there is statistical significance. This report looks at the South Dakota rates as compared to the U.S. national rates using SEER data.

Percent change: The difference between two rates expressed as a percentage.

Annual percent change (APC): The annual percent change is the average rate of change in a cancer rate per year in a given time frame indicating how fast or how slowly a cancer rate has increased or decreased each year over a period of years. A negative APC describes a decreasing trend, and a positive APC describes an increasing trend. In this report, a five-year period 1997-2001 was used and the calculations were made using *SEER STAT*.

Data source: All data, tables and figures come from the South Dakota Department of Health, *American Cancer Society Facts and Figures 2003* or *SEER Cancer Statistics Review 1975-2002*3and should be cited as such if taken out of this report in part. SEER data represents approximately 10% of the U.S population.

Disparity: Health disparity can be described in terms of “health inequality” or “health inequity.” It can be due to environment, acsess to/ utilization of health care, health status, and a particular health outcome that deserves scrutiny. Disparity can occur as a result of factors such as poverty, living in geographically underserved. Health disparities can be defined as a specific group bearing a disproportionate share of negative health outcomes compared to the general population, i.e., disease, disability, and death. areas and belonging to

specific minority groups. Health disparities in cancer are differences in the incidence, prevalence, mortality, and burden of diseases and other adverse health conditions that exist among specific population groups in the United States.

Limitations to Data Interpretation and Comparison

A number of factors need to be considered when reviewing cancer statistics and interpreting them. A cancer registry database is a fluid and dynamic database, therefore, the reported number of new cases in a particular race, gender and age-cancer category may change for the calendar year for which the data have already been reported in a previous publication. Additional cancer cases which have been previously overlooked for a given diagnosis year may be found and reported to the central registry. There may also be elimination of duplicate records for the same patient, often due to name changes or spelling corrections.

Rate comparisons: When comparing age-adjusted rates and age-specific rates based on fewer than 10 cases, rate comparisons are difficult to interpret. In comparing rates among geographic areas such as counties, states and health districts, the absolute numbers and differences in demographics should be considered, as well as clinical significance of the disease. Data quality indicators for each registry should also be reviewed. Interpretations without considering these factors may be misleading. There will also be differences between mortality statistics published by various agencies and the mortality rates in this report.

Racial misclassifications: When race is not specified in a source document and the default is to record these cases as white or unknown, the results are considered biased. Numerator error can occur because of misclassification.

Statistical significance: In South Dakota, counts can be very low; therefore, magnitude bias is inherent with confidence intervals and z-tests. For example, in year 2001, cervical cancer rates were 10 per 100,000 American Indian women with 2 deaths and 1.7 per 100,000 white women with 6 deaths, i.e., American Indian women had a cervical cancer age-adjusted rate six times higher than white women in South Dakota. However, the case counts were 2 for American Indians and 10 for whites. Small numbers result in wider confidence intervals, thus less confidence in the data.

Early detection/screening: Improved early detection/screening may produce increases in both incidence and survival rates. Increases may occur as a result of the introduction of new procedures. The interval between the time a cancer is diagnosed by a screening procedure and the time when it would have been diagnosed in the absence of screening procedures is called the lead-time. Changes in lead-time, for example, in breast cancer diagnosis, have led to an increased survival and a reduction of mortality.

Changes in diagnostic criteria: Early detection resulting from either screening or early response to symptoms may result in increasing diagnosis in small tumors that are not yet life-threatening. This may raise incidence and survival rates but without changes in mortality rates. Cancers likely to be affected are breast, colon, cervix uteri, prostate and melanoma. Prostate cancer is particularly prone to changing diagnostic criteria.

Staging: Advancement in diagnostic procedures may change in due time. Advances increase the probability that a given cancer will be diagnosed in a more advanced stage, for example, with new scanning methods, metastases can be detected. Therefore, if someone was previously diagnosed with a localized tumor, they may now be staged as distant. This is called stage migration and can affect the analysis of all solid tumors.